USING AGE AND SPATIAL FLOW STRUCTURES IN THE INDIRECT ESTIMATION OF MIGRATION STREAMS*

JAMES RAYMER AND ANDREI ROGERS

This article outlines a formal model-based approach for inferring interregional age-specific migration streams in settings where such data are incomplete, inadequate, or unavailable. The estimation approach relies heavily on log-linear models, using them to impose some of the regularities exhibited by past age and spatial structures or to combine and borrow information drawn from other sources. The approach is illustrated using data from the 1990 and 2000 U.S. and Mexico censuses.

Demographic estimation in countries with inadequate, inaccurate, or incomplete data-reporting systems often must rely on methods that are said to be "indirect." Such methods utilize inferential techniques that produce estimates of a particular variable by using data that may be only indirectly related to its value. The indirect estimation of fertility and mortality has a long history in demography. A common strategy there has been to combine empirical regularities with other information to fill in the missing data. Functional representations (Heligman and Pollard 1980) and relational representations (Brass 1974) of observed age patterns have occupied a central position in such efforts.

A somewhat dated 1983 United Nations manual serves as a useful entry into the vast literature on the topic. Unfortunately, like most of that literature, it ignores migration: "There are other demographic processes affecting the populations of these countries (migration, for example) which are not treated here" (United Nations 1983:1). More recently, a chapter on indirect estimation methods in an important text on formal demography (Preston, Heuveline, and Guillot 2000) also totally ignores migration. Demographic texts that do include topics on migration estimation tend to focus on residual methods (e.g., Rowland 2003; Siegel and Swanson 2004) similar to those presented by Bogue (1969:758–59) over 30 years ago.

The indirect estimation of migration flows has a briefer history, in part because the estimation task is more complicated. The age pattern of migrants depends on the directions of migration. To be effective, therefore, a method must somehow integrate the age pattern with the corresponding spatial pattern. Nonetheless, efforts to indirectly estimate migration *streams* continue (Ahmed and Robinson 1994; Hill 1985; Nair 1985; Schmertmann 1992; Warren and Kraly 1985; Warren and Peck 1980; Willekens 1999; Zaba 1987), notably those attempting to infer international or undocumented flows. This article adds to that literature contributing an operational method for estimating age- and origin-destination-specific migration flows from data on population stocks and auxiliary information. Much of the background for this approach comes from developments in spatial interaction modeling made by geographers in the late 1970s and early 1980s (Plane 1981, 1982; Snickars and Weibull 1977; Willekens 1980, 1982, 1983).

Two procedures are outlined, each of which requires a particular data set. One, *past migration* estimation, requires a complete set of migration flow data for one period and regional population stocks or gross migration flows for another. This method essentially

*James Raymer, Division of Social Statistics, School of Social Sciences, University of Southampton, Southampton, SO17 1BJ, United Kingdom; E-mail: raymer@soton.ac.uk. Andrei Rogers, Population Program, Institute of Behavioral Sciences, University of Colorado, Boulder. This work was supported by the National Institute of Child Health and Human Development and the National Science Foundation. The authors are grateful to Virgilio Partida Bush for providing the data on Mexico, to Rick Rogers for his comments and suggestions on an earlier draft of this article, and to the three anonymous reviewers who suggested useful revisions.

"updates" the migration data of a census in order to satisfy the marginal totals obtained or estimated for a later period of interest. If such migration data are not available, then the second method, *infant migration* estimation, instead uses an inferred migration spatial structure based on the birthplace-specific stock of children under 5 years of age at the time of the census.

This article sets out a methodology that allows for such an integration of estimation strategies. Since the problem is to predict the *number* of migrants by origin, destination, and age, the appropriate model is the log-linear model. It becomes a vehicle for determining whether the distribution of counts presented in the cells of a table matrix can be accounted for by an underlying structure. If the data are incomplete, then the underlying structure is determined by whatever auxiliary data are available, with the parameters of the log-linear model identifying the contributions of the various partial data sets to the predicted migration flows.

We begin the article with the description of a general modeling framework for describing and analyzing the age and spatial structures of interregional migration flows and show how it can be used to represent a particular pattern of age and spatial profiles. The approach decomposes an observed pattern into multiplicative components and then transforms that mathematical representation of migration into a statistical one by adopting the log-linear modeling framework for analyzing contingency tables. Two applications follow: the first is a discussion of past migration estimation, and the second is a discussion of infant migration estimation. A nine-region representation of migration flows in the United States and a four-region representation of migration flows in Mexico are used to illustrate the methods.

The results of this study should be of interest to at least two user communities: (1) migration analysts studying mobility patterns in data-poor, less-developed countries, and (2) population researchers faced with the prospective loss of the detailed migration data formerly contained in the "long-form" questionnaire of past U.S. decennial censuses and replaced in the forthcoming 2010 census by the smaller continuous monthly sampling survey called the American Community Survey.

DESCRIBING AND ESTIMATING THE AGE AND SPATIAL STRUCTURES OF INTERREGIONAL MIGRATION STREAMS

Migration flow patterns exhibit strong age and spatial regularities. In a discussion of new "laws" of migration, Tobler (1995:335) argued that "one of the most studied regularities is the age profile of migrants." He then focused on spatial patterns of migrants, presenting a table that "shows the correlation between all six U.S. state-to-state tables for the contiguous United States. Thirty-eight percent of the 1985–1990 migration table . . . can be explained by the 1935–1940 table, and 52% of it can be explained by the 1975–1980 table" (p. 336–37). A deeper analytical examination of this issue appears in a sequence of recent papers offering a formal definition of what constitutes the age and spatial structures of migration and how they can be represented by a multiplicative log-linear modeling framework (Raymer, Bonaguidi, and Valentini 2006; Rogers, Willekens, and Raymer 2001, 2002, 2003; Rogers, Willekens, Little, and Raymer 2002). This article adds to that research in two ways. First, a multiplicative component model is used to describe and model age-specific interregional migration flows in the United States and Mexico—two seemingly different situations. And, second, a consistent model-based framework is applied to estimate migration patterns using two types of auxiliary information, past migration and infant migration.

A Multiplicative Component Approach

Interregional migration flows (without age) can be disaggregated into four separate components (Rogers, Willekens, Little, and Raymer 2002): an *overall* component representing the level of migration, an *origin* component representing the relative "pushes" from each region, a *destination* component representing the relative "pulls" to each region, and a two-way *origin-destination interaction* component representing the impacts of physical or social distance between places (those not explained by the overall and main effects). This breakdown is multiplicative, such that

$$n_{ij} = (T)(O_i)(D_j)(OD_{ij}), \tag{1}$$

where n_{ij} is an observed flow of migration from region i to region j, T is the total number of migrants (i.e., n_{++}), O_i is the proportion of all migrants leaving from region i (i.e., n_{i+} / n_{++}), and D_j is the proportion of all migrants moving to region j (i.e., n_{+j} / n_{++}). The interaction component OD_{ij} is defined as n_{ij} / $[(T)(O_i)(D_j)]$, or the ratio of observed migration to expected migration (for the case of no interaction). This general type of model is called a *multiplicative component model*.

Next, consider the representation of *age-specific* migration patterns between these regions. The multiplicative component model for this table is specified as

$$n_{ijx} = (T)(O_i)(D_j)(A_x)(OD_{ij})(OA_{ix})(DA_{jx})(ODA_{ijx}), \tag{2}$$

where A_x is the proportion of all migrants in age group x. This model is more complicated because there are now three two-way interaction components and a single three-way interaction component between the origin, destination, and age variables. However, the interpretations of the parameters remain relatively simple and follow the same format as presented for the two-way table. That is, the interaction components represent ratios of observed flows or marginal totals to expected ones. For example, the destination-age interaction (DA_{jx}) component is calculated as n_{+jx} / $[(T)(D_j)(A_x)]$ and represents the ratios of observed age patterns of in-migration to each region divided by the expected age pattern of in-migration.

The Log-Linear Model

The multiplicative component descriptive model set out in Eq. (2) can be expressed as a *saturated* log-linear statistical model,

$$\ln(n_{ijx}) = \lambda + \lambda_i^O + \lambda_j^D + \lambda_x^A + \lambda_{ij}^{OD} + \lambda_{ix}^{OA} + \lambda_{jx}^{DA} + \lambda_{ijx}^{ODA}, \tag{3}$$

where the λs are simply the natural logarithms of the variables appearing in Eq. (2). In multiplicative form, this model is expressed as

$$n_{ijx} = \tau\tau_i^O\tau_j^D\tau_x^A\tau_{ij}^{OD}\tau_{ix}^{OA}\tau_{jx}^{DA}\tau_{ijx}^{ODA}, \tag{4}$$

where the τs denote the model's multiplicative parameters or "effects." We use this form to be consistent with the multiplicative component model. The saturated model is expressed as (ODA), using the notation set out in Agresti (2002:320). The parameters of the log-linear model can be analyzed by using standard statistical techniques for categorical data analysis to identify key structures in the data.

Reduced forms of the models set out in Eqs. (3) and (4) are called *unsaturated* models. For example, the model that includes only the *main effects* of origin, destination, and age is specified as

$$\hat{n}_{ijx} = \tau\tau_i^O\tau_j^D\tau_x^A, \tag{5}$$

where $\hat{n}_{ijx}$ denotes the predicted age-specific migration flows. This model assumes independence between each of the categories of origin, destination, and age and is designated (O, D, A). A model that includes the interaction between origin and destination plus all of the main effects is designated as (OD, A) and is specified as $\hat{n}_{ijx} = \tau\tau_i^O\tau_j^D\tau_x^A\tau_{ij}^{OD}$. Such notations are used because these models are hierarchical; that is, for two-way interaction terms, the main effect parameters must be included, and for three-way interaction terms, all the main effects and two-way interactions must be included.

Migration flow tables are complicated because they can mix migrants with nonmigrants or intraregional migrants. To remove nonmigrant elements from the analysis, structural zeros can be inserted using an indicator function (Agresti 2002; Willekens 1983). When structural zeros are *included* in the model, Eq. (5) is called a *quasi-independence* model.

An *offset*, a matrix with auxiliary information, can be used to incorporate such information (as well as structural zeros) to improve the estimation procedure. Auxiliary information can be, for example, a historical table of migration flows. The log-linear-with-offset model is specified as

$$\hat{n}_{ijx} = n_{ijx}^{*}\tau\tau_i^O\tau_j^D\tau_x^A, \tag{6}$$

where n_{ijx}^{*} denotes the auxiliary information (refer to Rogers et al. 2003:60–61). In this case, the flows contained in the offset would be forced to fit the marginal totals represented by the overall level and main effects of age, origin, and destination.

We use known data in this article to test our ideas. The migrant-only models make the strong assumption that the current marginal totals are known—that is, the overall level of migration, the proportions migrating from and to each region, and the proportions in each age group are given. Our emphasis is on identifying and modeling the age and spatial patterns within these marginal totals. However, some examples are provided in the past migration and infant migration estimation sections that do not make such a strong assumption and instead use age-specific population stocks at the beginning and end of the interval as the marginal total information to estimate both migrants and nonmigrants. Of course, the marginal totals could also have been modeled independently (Little and Rogers 2007). Furthermore, the modeling framework presented in this paper can be applied to unknown situations. For example, the multiplicative component approach has been applied to project future age-specific interregional migration flows in Italy (Raymer et al. 2006) and to estimate age-specific international flows between countries in the European Union, Iceland, Norway, and Switzerland during the 2001–2002 period (Raymer forthcoming).

The models in this paper are evaluated using the likelihood ratio statistic, G^2,

$$G^2 = 2\Sigma n_{ijx} \ln(n_{ijx} / \hat{n}_{ijx}), \tag{7}$$

where values closest to zero are associated with "good" fits (see, e.g., Agresti 2002). We also use the coefficient of determination, R^2, when examining particular age-specific flow estimates. The former is useful for assessing overall fit in terms of levels; the latter is useful for assessing overall fit in terms of patterns (or shapes).

APPLICATIONS

The age and spatial structures of interregional migration in the United States and Mexico during the 1995–2000 period may be described by using the multiplicative components model set out above. Such an analysis follows a hierarchical format, starting with the overall level component and ending with the three two-way interaction components. The three-way interactions between origin, destination, and age are not analyzed for two reasons. First, most of the structure found in the migration patterns is captured by the overall, main, and two-way interaction effects. Second, although patterns are often found in the three-way interactions, it is tedious to incorporate them into the modeling process, and their interpretation is more difficult. Therefore, we shall just focus on the simpler and more powerful aspects of the model represented by the other seven terms found in Eq. (2).

The United States

To illustrate the advantages of analyzing migration in terms of multiplicative components, consider the U.S.-born migration flows between the nine Census Bureau–defined regions (Divisions) during the 1995–2000 time period set out in Panel A of Table 1. Note that non-migrants (i.e., n_{ii}) are not included in the table. During this period, 14.6 million U.S.-born persons over the age of 5 years made an interregional migration. Nearly half of all migrants came from the East North Central, South Atlantic, and Pacific regions, and about a quarter of all migrants went to the South Atlantic region. The largest origin-destination-specific flow was from the Middle Atlantic region to the South Atlantic region.

The multiplicative components corresponding to the migration flows discussed above are set out in Panel B of Table 1. Note that the overall component (T) is set out in the total sum (i.e., n_{++}) location of the table, the origin components (O_i) are set out in the row-sum locations (i.e., n_{i+}), the destination components (D_j) are set out in the column-sum locations (i.e., n_{+j}), and the origin-destination interaction components (OD_{ij}) are set out in the cells inside the marginal totals (i.e., n_{ij}). For example, consider the Middle Atlantic to South Atlantic flow of 1,084 thousand persons disaggregated into the four multiplicative components:

$$n_{25} = (T)(O_2)(D_5)(OD_{25})$$

$$= n_{++}\left(\frac{n_{2+}}{n_{++}}\right)\left(\frac{n_{+5}}{n_{++}}\right)\left[\frac{n_{25}}{(n_{++})\left(\frac{n_{2+}}{n_{++}}\right)\left(\frac{n_{+5}}{n_{++}}\right)}\right]$$

$$= (14{,}657)\left(\frac{2{,}097}{14{,}657}\right)\left(\frac{3{,}573}{14{,}657}\right)\left(\frac{1{,}084}{511}\right)$$

$$= (14{,}657)(0.143)(0.244)(2.120)$$

$$= 1{,}084,$$

where the subscripts 2 and 5 denote the Middle Atlantic and South Atlantic regions, respectively. The interpretations of these components are relatively simple. The overall component is the reported total number of U.S.-born interregional migrants aged 5 years and over; 14.6 million persons made an interregional move between 1995 and 2000. The origin component represents the shares of all migrants from each region: 14% of all migrants originated in the Middle Atlantic region. The destination component represents the shares of all migrants to each region: 24% of all migrants moved to the South Atlantic region. And, finally, the interaction component represents the ratio of observed migration to expected migration; there were roughly two observed migrants for every one expected migrant. The expected flow is based on the marginal total information, for example, $(T)(O_2)(D_5)$.

The ratios of observed to expected flows set out in Panel B of Table 1 capture the relative association or "interaction" between regions, so, for example, the interaction component value of 2.12 indicates a strong association between the Middle Atlantic and South Atlantic regions. Other flows that exhibited high levels of association (over 2.0) were New England–Middle Atlantic, Middle Atlantic–New England, West North Central–East North Central, South Atlantic–Middle Atlantic, South Atlantic–East South Central, Mountain-Pacific, and Pacific-Mountain. In all of these cases, the regions share borders with each other.

Table 1. The Spatial Structure of Interregional Migration in the United States (in thousands), 1995–2000

Origin	Destination: New England	Middle Atlantic	East North Central	West North Central	South Atlantic	East South Central	West South Central	Mountain	Pacific	Total
A. Observed Flows										
New England	0	167	61	22	298	23	41	59	100	771
Middle Atlantic	245	0	199	54	1,084	74	105	145	191	2,097
East North Central	68	161	0	297	674	280	223	273	241	2,217
West North Central	25	48	270	0	185	63	205	215	145	1,157
South Atlantic	168	437	413	139	0	393	314	215	301	2,380
East South Central	18	40	185	47	379	0	159	54	67	947
West South Central	37	76	184	188	358	179	0	235	226	1,482
Mountain	43	72	154	166	197	53	222	0	472	1,379
Pacific	92	151	230	180	397	101	310	766	0	2,227
Total	696	1,150	1,696	1,093	3,573	1,165	1,581	1,962	1,741	14,657
B. Multiplicative Components										
New England	0.000	2.755	0.686	0.388	1.583	0.374	0.494	0.573	1.092	0.053
Middle Atlantic	2.464	0.000	0.820	0.344	2.120	0.445	0.466	0.516	0.765	0.143
East North Central	0.647	0.926	0.000	1.794	1.247	1.589	0.934	0.920	0.913	0.151
West North Central	0.460	0.523	2.015	0.000	0.658	0.687	1.647	1.390	1.054	0.079
South Atlantic	1.486	2.341	1.500	0.786	0.000	2.075	1.225	0.674	1.063	0.162
East South Central	0.393	0.532	1.688	0.664	1.641	0.000	1.554	0.425	0.592	0.065
West South Central	0.519	0.652	1.071	1.704	0.992	1.516	0.000	1.185	1.281	0.101
Mountain	0.653	0.662	0.967	1.612	0.587	0.481	1.495	0.000	2.883	0.094
Pacific	0.873	0.862	0.892	1.082	0.732	0.570	1.291	2.570	0.000	0.152
Total	0.047	0.078	0.116	0.075	0.244	0.080	0.108	0.134	0.119	14,657

The extension of the above analysis to include age is straightforward. The age groups used in this article start with 5–9 years and end with 85+ years and are measured at the time of the census. There are 17 age groups total. The age main effect component describes the age composition of all migrants in the multiregional system. The origin-age interaction components can be used to identify important differences between age-specific out-migration levels from each region and the overall age profile of migration found in the corresponding expected flows (i.e., $(T)(O_i)(A_x)$). The same is true for the destination-age interaction components, but with a focus on the differences between age-specific in-migration levels to each region and their corresponding expected flows (i.e., $(T)(D_j)(A_x)$).

The origin-age and destination-age interaction components are useful for identifying relative differences found in age patterns of in-migration and out-migration, respectively. For example, in examining the origin-age components (not shown for space reasons; see Figure 3 for example), we found particularly high propensities of young adult migration from the New England, Middle Atlantic, East North Central, and West North Central regions. The opposite was true for young adults migrating from the South Atlantic and Pacific regions. Not surprisingly, out-migration from the New England, Middle Atlantic, and East North Central regions contained age profiles with higher than expected levels around retirement years. The same was true for migration to the South Atlantic and Mountain regions found in the destination-age interaction components (also not shown; see Figure 3 for example).

Finally, we compare *unsaturated* log-linear models to analyze underlying structures in the U.S. migration data. All models include structural zeros to remove nonmigrants from the predictions, and the results, set out in Table 2, are compared using the likelihood ratio statistic. The most obvious finding is that the origin-destination interaction term is very important for accurately predicting the age-specific migration flows. Most of the flows do not contain a large retirement peak or major deviations from the overall age profile of migration. However, the fits are slightly improved when the origin-age or destination-age interactions (with the latter doing a better job) are included. Of course, to capture different age profiles found in some of the flows, such as those with retirement peaks, origin-age or destination-age interactions have to be included.

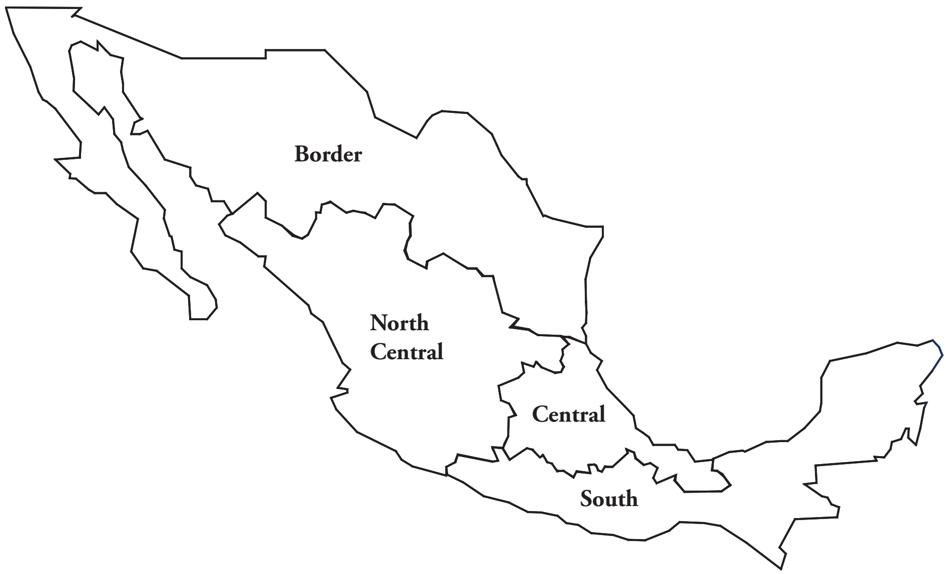
Mexico

The Mexican interregional migration data come from the 1990 and 2000 censuses and represent persons born in Mexico. The country has been divided into four regions on the

Table 2. Unsaturated Log-Linear Model Fits: Age-Specific Interregional Migration Flows in the United States, 1995–2000

Model	Likelihood Ratio Statistic, G^2	*df*	G^2 / df
(*O*, *D*, *A*)	4,068,146	1,191	3,416
(*OD*, *A*)	545,855	1,136	481
(*OA*, *D*)	3,909,131	1,063	3,677
(*DA*, *O*)	3,817,146	1,063	3,591
(*OD*, *OA*)	386,839	1,008	384
(*OD*, *DA*)	294,855	1,008	293
(*OA*, *DA*)	3,678,153	935	3,934
(*OD*, *OA*, *DA*)	163,392	880	186

Figure 1. The Four Regions of Mexico



basis of economics and history (see Figure 1). The Border region has the most formal employment. The North Central region is an area of medium-level development, with an economy focused on manufacturing and export agriculture. The Central region, formerly the most dynamic area in Mexico, remains the country's financial and political hub, with Mexico City, the capital, as its center. The South region, historically the country's poorest, currently has an economy based on tourism and petroleum.

The aggregate migration flows between the Border, North Central, Central, and South regions during the 1995–2000 period are set out in Panel A of Table 3, and the corresponding multiplicative components are presented in Panel B. During the 1995–2000 period, 1.76 million persons over the age of 5 years (in the year 2000) made an interregional migration in Mexico, with 40% coming from the Central region. The Border region *received* nearly the same number of migrants. The largest origin-destination-specific flow was the North Central–Border flow. This flow had an origin-destination association of 1.7. Other flows that exhibited high levels of association (i.e., values greater than 1.5) were the Border–North Central, Central-South, and the South-Central flows. In all of these instances, the regions share a border.

Some interesting regional patterns were found in the origin-age and destination-age patterns (again, not shown for space reasons; see Figure 6 for examples). For example, children exhibited higher than expected levels of migration from the Border region. Young adults had higher than expected levels of migration from the South region, whereas from the Border region, the flows were lower than expected. Persons older than 25 years were less likely to leave the South, whereas persons aged 30–44 were more likely to leave the Border region. And the elderly were more likely to leave the North Central region. As for the destination-age interaction components, young adults exhibited higher than expected levels of migration to the Border and Central regions. Elderly migrants clearly

Table 3. The Spatial Structure of Interregional Migration in Mexico, 1995–2000

	Destination				
Origin	Border	North Central	Central	South	Total
A. Observed Flows					
Border	0	122,915	69,709	20,883	213,507
North Central	308,712	0	134,961	28,589	472,262
Central	278,185	219,251	0	199,803	697,239
South	89,973	89,041	201,156	0	380,170
Total	676,870	431,207	405,826	249,275	1,763,178
B. Multiplicative Components					
Border	0.000	2.354	1.419	0.692	0.121
North Central	1.703	0.000	1.242	0.428	0.268
Central	1.039	1.286	0.000	2.027	0.395
South	0.616	0.958	2.299	0.000	0.216
Total	0.384	0.245	0.230	0.141	1,763,178

Note: Numbers refer to Mexican-born persons.

preferred the Central region (home to the capital, Mexico City, and its healthcare facilities) to other regions.

A log-linear analysis was also carried out for the Mexican flow data. Again, the origin-destination interaction term was found to be very important for accurately predicting the age-specific migration flows. Most of the age-specific regional flows did not deviate much from the overall age profile of migration. However, the fits were slightly improved when the origin-age or destination-age interactions were included (with the latter doing a better job).

PAST MIGRATION ESTIMATION

The 1995–2000 age-specific interregional migration patterns in the United States and in Mexico are estimated in this section, using some of the structures found in the previous census. In particular, the log-linear-with-offset model (i.e., Eq. (6)) is applied to estimate the 1995–2000 age-specific interregional migration flows. The offset in this case is the matrix of observed 1985–1990 age-specific interregional migration flows. Depending on the available data, the estimation can focus on (1) migrants or on (2) both migrants and nonmigrants. The first implies that the aggregate numbers of persons in-migrating and out-migrating for each region are known, whereas the second implies that only the beginning and ending regional population stocks are known (a more common situation). For the second case, T denotes the overall population size of persons aged 5 years and older, O_i denotes the proportion of the population residing in a region at the beginning of the interval, D_j denotes the proportion of the population residing in a region at the end of the interval, and A_x denotes the proportions of the total population in each age group x. The main concern with modeling both migrants and nonmigrants is the tendency of nonmigrants to dominate the results. During the 1985–1990 and 1995–2000 periods, about 93% of the U.S.-born populations and about 98% of the Mexican-born populations were nonmigrants. For direct estimation modeling, this means that any substantial changes in the nonmigrant origin-destination interaction components will have a sizable impact on the predicted flows of migration. To demonstrate the implications for the U.S. and Mexico migration estimations, we used two offsets to estimate the 1995–2000 age-specific interregional migration flows: one that included only migrants and another that included both migrants and nonmigrants.

The United States

The age and spatial structures of U.S. interregional migration have exhibited stability over time. The age main effect components for the 1985–1990 and 1995–2000 periods are set out in Figure 2. The main differences between the two periods are that the labor force peak became slightly wider in the later period and that the retirement peak disappeared entirely. New England's and South Atlantic's origin-age and destination-age interaction components have been set out in Figure 3 for the two migration periods as a another example of stability over time. Here, the most noticeable differences were found in the retirement years, where the patterns of the 1995–2000 period were less extreme than in the 1985–1990 period. Overall, the comparisons of the age and spatial structures of migration between the two periods show continuity over time and suggest that a model relying on the 1990 census data to estimate the 1995–2000 migration patterns should perform well.

The log-linear-with-offset model was applied to estimate the 1995–2000 age-specific interregional migration flows by "borrowing" the two-way and three-way associations found in the migration data captured in the previous census. Both the "migrants only" and "with nonmigrants" models performed well, as illustrated with some selected flows in Figure 4. In particular, the migrants-only R^2 values were 0.987, 0.973, 0.994, and 0.971 for the New England–Middle Atlantic, Middle Atlantic–South Atlantic, South Atlantic–Middle

Figure 2. The Age Main Effect Components of Interregional Migration in the United States, 1985–1990 and 1995–2000

Figure 3. The Origin-Age and Destination-Age Interaction Components of Interregional Migration in the United States, 1985–1990 and 1995–2000: New England and South Atlantic Regions

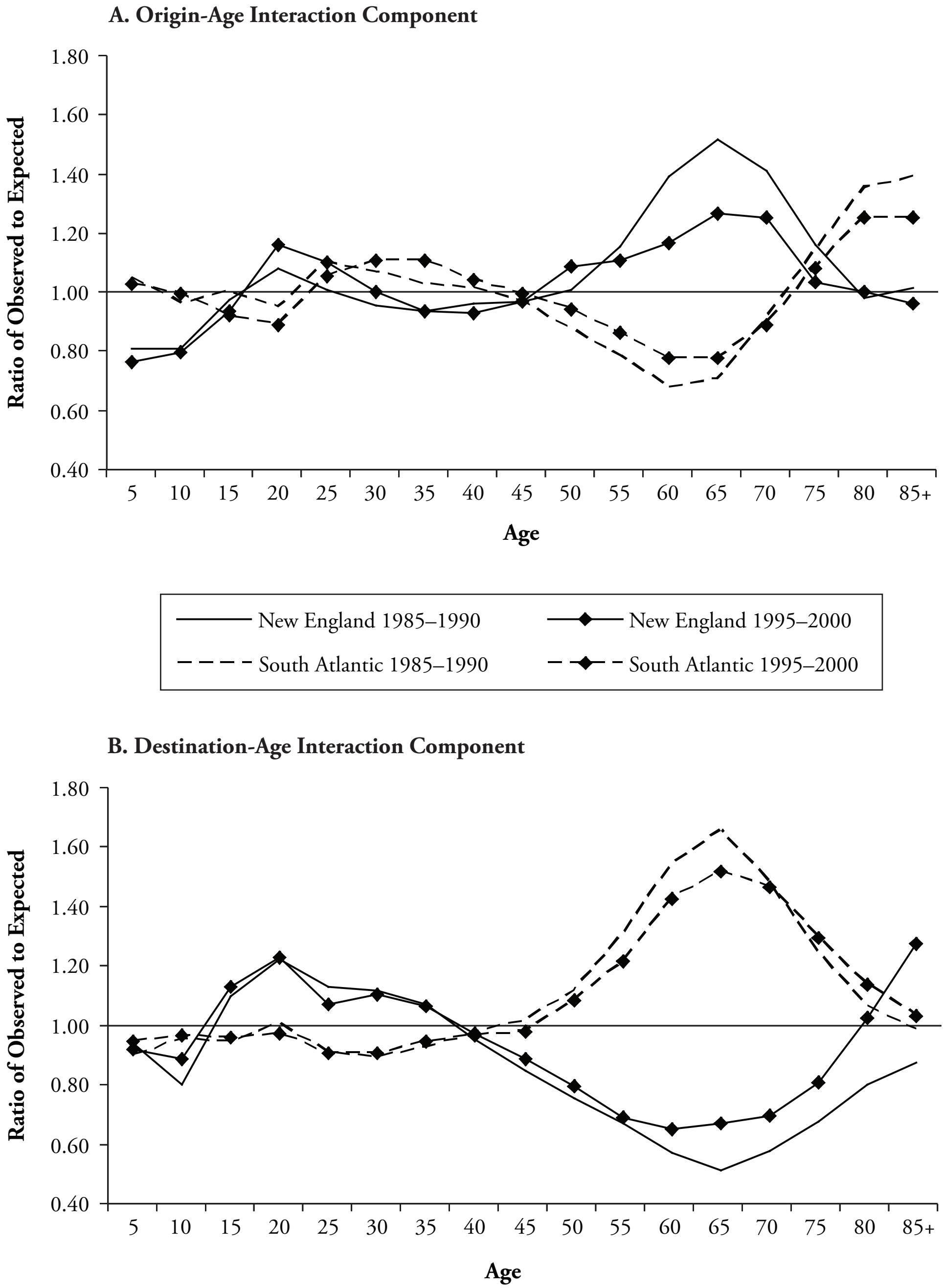
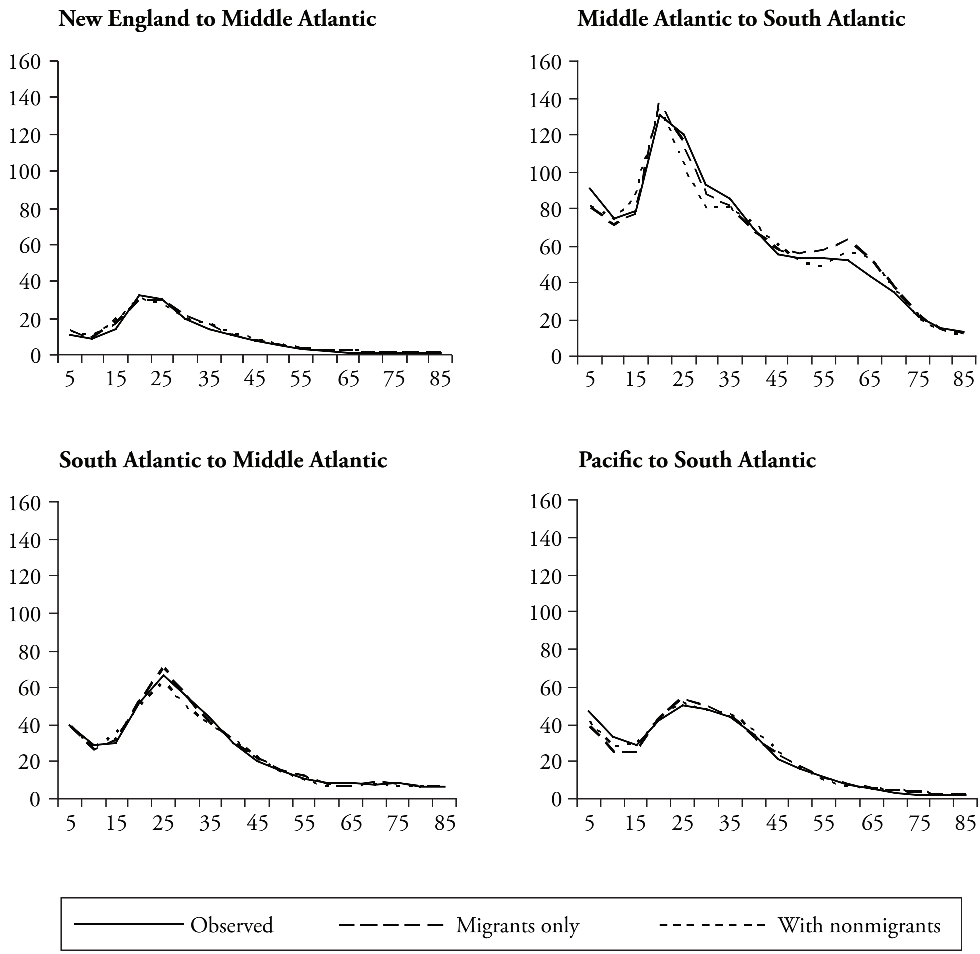


Figure 4. A Comparison of Past Migration Log-Linear Model Predictions: Selected Age-Specific Interregional Migration Flows (in thousands) in the United States, 1995–2000



Atlantic, and Pacific–South Atlantic flows, respectively. For the migrants and nonmigrants model, the R^2 values were 0.971, 0.958, 0.986, and 0.984, respectively. The G^2 statistics, calculated for these flows, correspond with the R^2 patterns. Note that these flows were selected because of their different age-specific shapes. The South Atlantic–Middle Atlantic flow has a relatively sharp labor force peak in comparison to the relatively flat-peaked New England–Middle Atlantic flow and the wide labor force curved Pacific–South Atlantic flow. The Middle Atlantic–South Atlantic flow is an example of a flow with a retirement peak. In terms of overall fit, the migrant-only model performed better with a G^2 of 236,326 versus –425,830 for the migrants and nonmigrants model.

Table 4. A Comparison of Mexico's Migration Spatial Structures Over Time: Ratios of 1995–2000 Multiplicative Components to 1985–1990 Multiplicative Components

	Destination				
Origin	Border	North Central	Central	South	Total
Border		1.068	1.128	1.273	1.083
North Central	0.857		1.086	1.297	0.886
Central	1.240	0.831		0.942	1.014
South	1.674	1.253	0.876		1.102
Total	1.148	0.889	0.906	1.036	1.056

Mexico

A comparison of Mexico's spatial structures of migration over time illustrates some interesting patterns (Table 4). The overall level increased by 6%. The share of migration originating in the Border and South regions increased by 8% and 10%, respectively, and decreased by slightly more than 11% in the North Central region. The proportions of migrants going to the Border region increased substantially, whereas those going to the North Central and Central regions declined. For the origin-destination associations, the extremes were those corresponding with the South to Border flow, which increased by 67%, and the Central to North Central flow, which decreased by 17%.

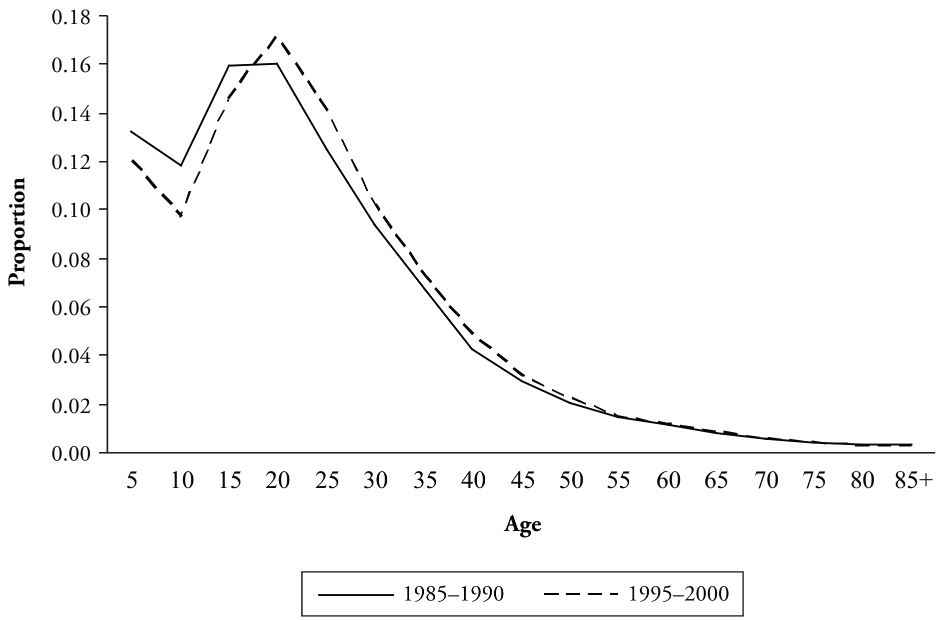
The age main effect components for the 1985–1990 and 1995–2000 periods are set out in Figure 5. The main difference between the two periods is that the labor force peak shifted slightly to the right and that there were slightly lower proportions of young children migrating in the later period. A comparison of origin-age and destination-age interaction components for the Border and South regions during the two migration periods shows strong continuity over time, as illustrated in Figure 6. These age profiles illustrate that young adults were more likely to migrate from the South region and to migrate to the Border region. Not surprisingly, young adults were also less likely to migrate from the Border region and migrate to the South region. Again, the comparisons of the age and spatial structures of migration between the two periods show general continuity over time and suggest that a model relying on the 1990 census data to estimate the 1995–2000 migration patterns should perform well.

As for the U.S. case study, two offsets were used to estimate the 1995–2000 age-specific interregional migration flows in Mexico: one that included only migrants and another that included both migrants and nonmigrants. The model that included both migrants and nonmigrants overpredicted the number of migrants by 249,000 and had a G^2 of –173,004. For the migrants-only model, the G^2 was 26,420. A selection of the estimated flows is presented in Figure 7. For the migrants-only model, the R^2 values were 0.986, 0.996, 0.999, and 0.998 for the Border–North Central, North Central–Border, Central-South, and South-Central flows, respectively, and 0.993, 0.996, 0.996, and 0.998, respectively, for the migrants and nonmigrants model. Note, for the above flows, that the G^2 statistics were all substantially closer to zero for the migrants-only model.

INFANT MIGRATION ESTIMATION

A new method for indirectly estimating migration patterns was recently put forward by Rogers and Jordan (2004), in which regional birthplace-specific population stock data of 0- to 4-year-olds was used to predict age-specific interregional patterns of migration in the

Figure 5. The Age Main Effect Components of Interregional Migration in Mexico, 1985–1990 and 1995–2000



United States. The first age group was used to capture the interregional patterns (not the levels) of the five-year interval migration question. That is, if a child is living in a different place than his or her place of birth, that child must have migrated at least once during the past five years. The same cannot be said for other age groups. And the reason why the migration pattern of a single age group can predict the corresponding patterns for other age groups comes from the age regularities found in observed migration patterns.

Migration propensities differ greatly according to age. Typically, an age-specific profile of migration shows a downward slope from the early childhood age groups to about age 16, is followed by a rise to a peak in the young adult age groups (usually around age 22), and then gradually tapers off to the oldest age groups. This "standard" age profile of migration can be fully described by using a multi-exponential model migration schedule (Rogers and Castro 1981; Rogers and Little 1994).

The most often used model migration schedule is the seven-parameter version:

$$N_{ijx} = a_0 + a_1 \exp\left(-\alpha_1 x\right) + a_2 \exp\left\{-\alpha_2\left(x-\mu_2\right) - \exp\left[-\lambda_2\left(x-\mu_2\right)\right]\right\}, \quad i \neq j, \tag{8}$$

where N_{ijx} denotes standardized (to unit area) age profiles of migration from region i to region j at age group x. The a_0, a_1, and a_2 are level parameters, whereas the α_1, α_2, μ_2, and λ_2 parameters are shape parameters. This schedule can be used, for example, to represent the aggregate age profiles of interregional migration (standardized to unit area) for the

Figure 6. The Origin-Age and Destination-Age Interaction Components of Interregional Migration in Mexico, 1985–1990 and 1995–2000: Border and South Regions

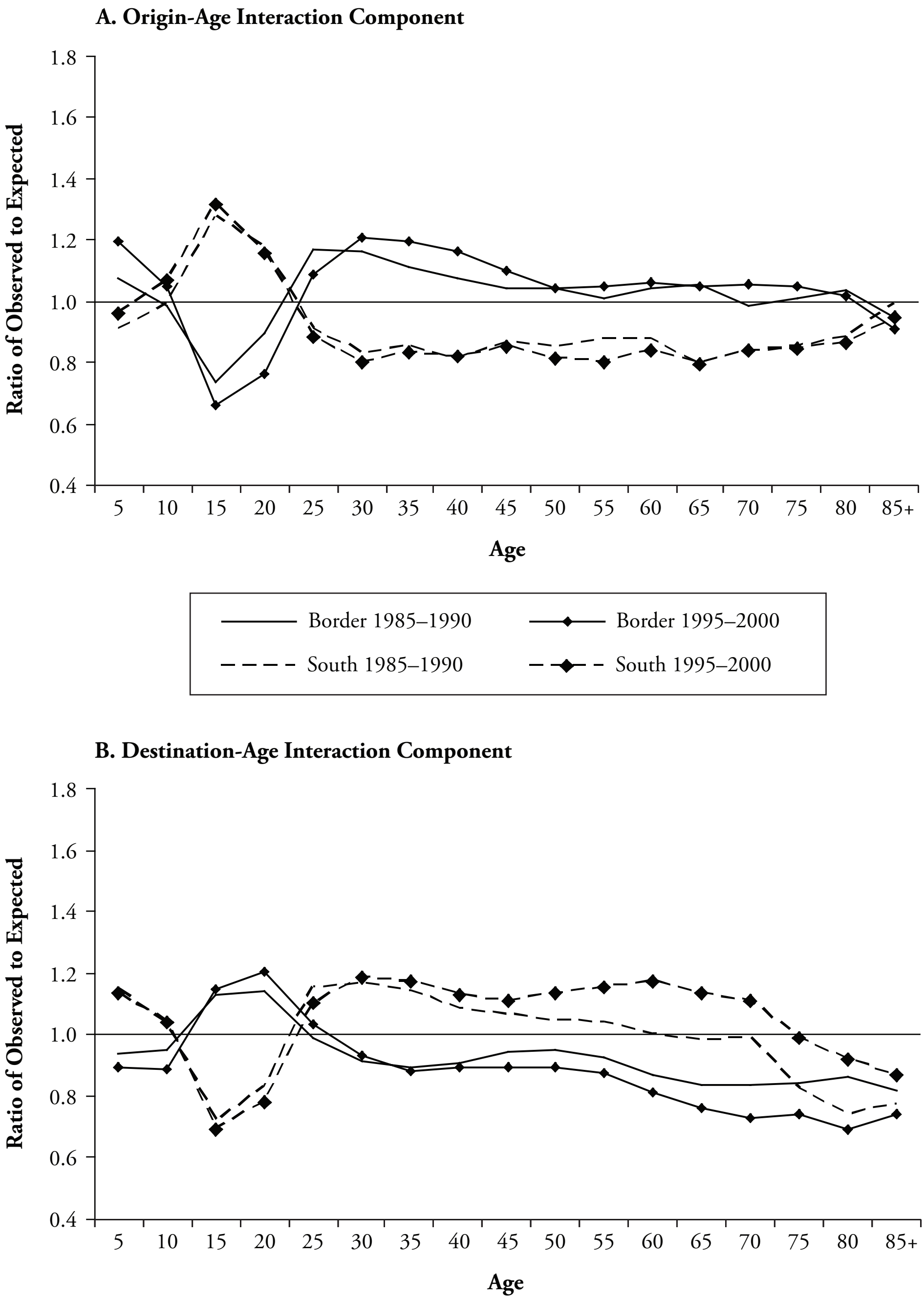
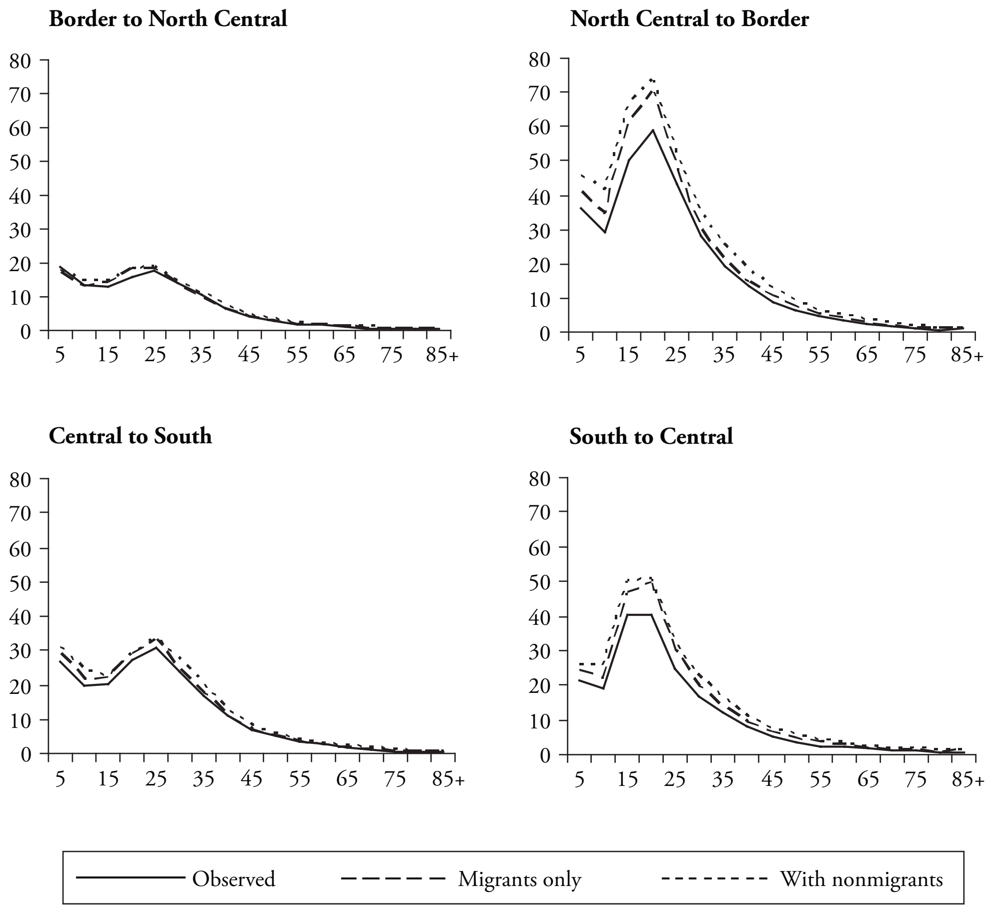


Figure 7. A Comparison of Past Migration Log-Linear Model Predictions: Selected Age-Specific Interregional Migration Flows (in thousands) in Mexico, 1995–2000



United States and Mexico during the 1995–2000 period. The estimated parameters for the United States are

$$\hat{N}_{ijx} = 0.00 + 0.11\exp(-0.04x) + 0.19\exp\{-0.061(x-19.24) - \exp[-0.23(x-19.24)]\}.$$

And for Mexico, they are

$$\hat{N}_{ijx} = 0.00 + 0.18\exp(-0.07x) + 0.22\exp\{-0.066(x-15.20) - \exp[-0.30(x-15.20)]\}.$$

When the estimated values associated with these curves are compared with the observed values, the R^2 values are 0.934 and 0.927, respectively.

The log-linear-with-offset model can be thought of as a relational model (Rogers et al. 2003). In this situation, the offset is the collection of 0- to 4-year-old birthplace-specific population stocks. We can specify a log-linear-with-offset model that uses the 0- to 4-year-old birthplace-specific population stocks to predict the aggregate patterns (assuming the marginal totals are known):

$$\hat{n}_{ij} = n^*_{ij} \nu \nu^O_i \nu^D_j , \tag{9}$$

where the offset n^*_{ij} contains the "migration" patterns of those aged 0–4 years at the time of the census, and effectively serves as a "proxy" for the interaction patterns of the current migration flows.

For age-specific patterns, the log-linear-with-offset model specified in Eq. (6) can be used. In this case, the offset contains structural zeros in the diagonal and the "migration" patterns of those aged 0–4 years at the time of the census in the off-diagonals. The overall age profile and aggregate proportions migrating from and to each region are assumed to be known. If instead one has to work with population totals, then one needs to estimate or borrow the aggregate age-specific proportions of migrants and nonmigrants. The model used in this case would be

$$\hat{n}_{ijxz} = n^*_{ijxz} \nu \nu^O_i \nu^D_j \nu^A_x \nu^M_z \nu^{AM}_{xz} , \tag{10}$$

where M denotes migrant status (i.e., migrant or nonmigrant status). This specification is required to distinguish between the age profiles of migrants and those of nonmigrants.

The United States

The 0- to 4-year-old "migration" patterns for U.S.-born persons are set out in Table 5. The spatial structure of these "infant" migrants closely resembles that of the period migrants set out earlier in Table 1. The predicted aggregate flows from New England and South Atlantic are presented in Figure 8. These predicted flows come from the model specified in Eq. (9), but with two alternative offsets being used: (1) migrants only and (2) migrants and nonmigrants. Although both models appear to predict the observed data well, the migrants-only model did considerably better. The likelihood ratio statistics for the two models were 132,799 and –1,632,755, respectively. The corresponding R^2 values were 0.985 and 0.955, respectively.

The *age-specific* predictions using the models in Eq. (6) and Eq. (10) also did well, capturing the levels and most of the age profiles. Examples of such predictions are set out in Figure 9. Our illustration applied a single age profile to estimate all age-specific patterns. The age profile is the same for both the migrants-only and the migrants and non-migrants models. This means that the shapes of some flows, such as the retirement migration peak found in the Middle Atlantic to South Atlantic flow, were not captured. For the flows set out in Figure 9, the R^2 values were 0.878, 0.940, 0.967, and 0.948 for the New England–Middle Atlantic, Middle Atlantic–South Atlantic, South Atlantic–Middle Atlantic, and Pacific–South Atlantic flows, respectively. The corresponding likelihood ratio statistics were lower for the migrants-only model, except for the Pacific–South Atlantic flow. Overall, the migrants-only model performed better with an overall G^2 of 678,641 versus 890,321 for the migrants and nonmigrants model.

Mexico

The input data for the indirect estimation of age-specific migration flows between regions in Mexico are set out in Table 6. Examples of the age-specific predictions (which are the same as those used in the previous section on past migration estimation) are set out in Figure 10. Although both models appear to predict the observed data well, the

Table 5. The Spatial Structure of 0- to 4-Year-Old Birthplace-Specific Population Stocks in the United States, 2000

	Destination									
Origin	New England	Middle Atlantic	East North Central	West North Central	South Atlantic	East South Central	West South Central	Mountain	Pacific	Total
A. Observed Flows (in thousands)										
New England	0	14	5	2	17	2	3	3	6	52
Middle Atlantic	17	0	19	5	62	6	10	8	14	140
East North Central	5	15	0	25	39	24	17	15	18	158
West North Central	2	5	36	0	15	6	19	14	11	106
South Atlantic	12	42	44	13	0	35	30	15	28	219
East South Central	2	5	27	6	34	0	18	4	7	103
West South Central	4	8	20	19	33	15	0	23	26	147
Mountain	4	8	13	14	17	5	20	0	43	123
Pacific	8	15	25	18	42	10	32	65	0	214
Total	54	111	189	102	258	102	148	147	152	1,264
B. Multiplicative Components										
New England	0.000	3.126	0.678	0.500	1.572	0.464	0.505	0.469	0.915	0.041
Middle Atlantic	2.824	0.000	0.886	0.472	2.152	0.536	0.594	0.467	0.843	0.111
East North Central	0.708	1.063	0.000	1.918	1.204	1.903	0.929	0.835	0.958	0.125
West North Central	0.471	0.501	2.242	0.000	0.681	0.674	1.502	1.096	0.867	0.084
South Atlantic	1.317	2.166	1.344	0.743	0.000	1.974	1.160	0.594	1.063	0.173
East South Central	0.506	0.506	1.756	0.735	1.626	0.000	1.450	0.373	0.550	0.082
West South Central	0.587	0.653	0.920	1.591	1.099	1.231	0.000	1.328	1.438	0.117
Mountain	0.729	0.733	0.726	1.438	0.657	0.475	1.380	0.000	2.862	0.098
Pacific	0.882	0.779	0.771	1.017	0.967	0.559	1.277	2.617	0.000	0.170
Total	0.043	0.088	0.149	0.081	0.204	0.080	0.117	0.116	0.121	1,264

Figure 8. A Comparison of Infant Migration Log-Linear Model Predictions: Interregional Migration Flows (in thousands) From New England and South Atlantic, 1995–2000

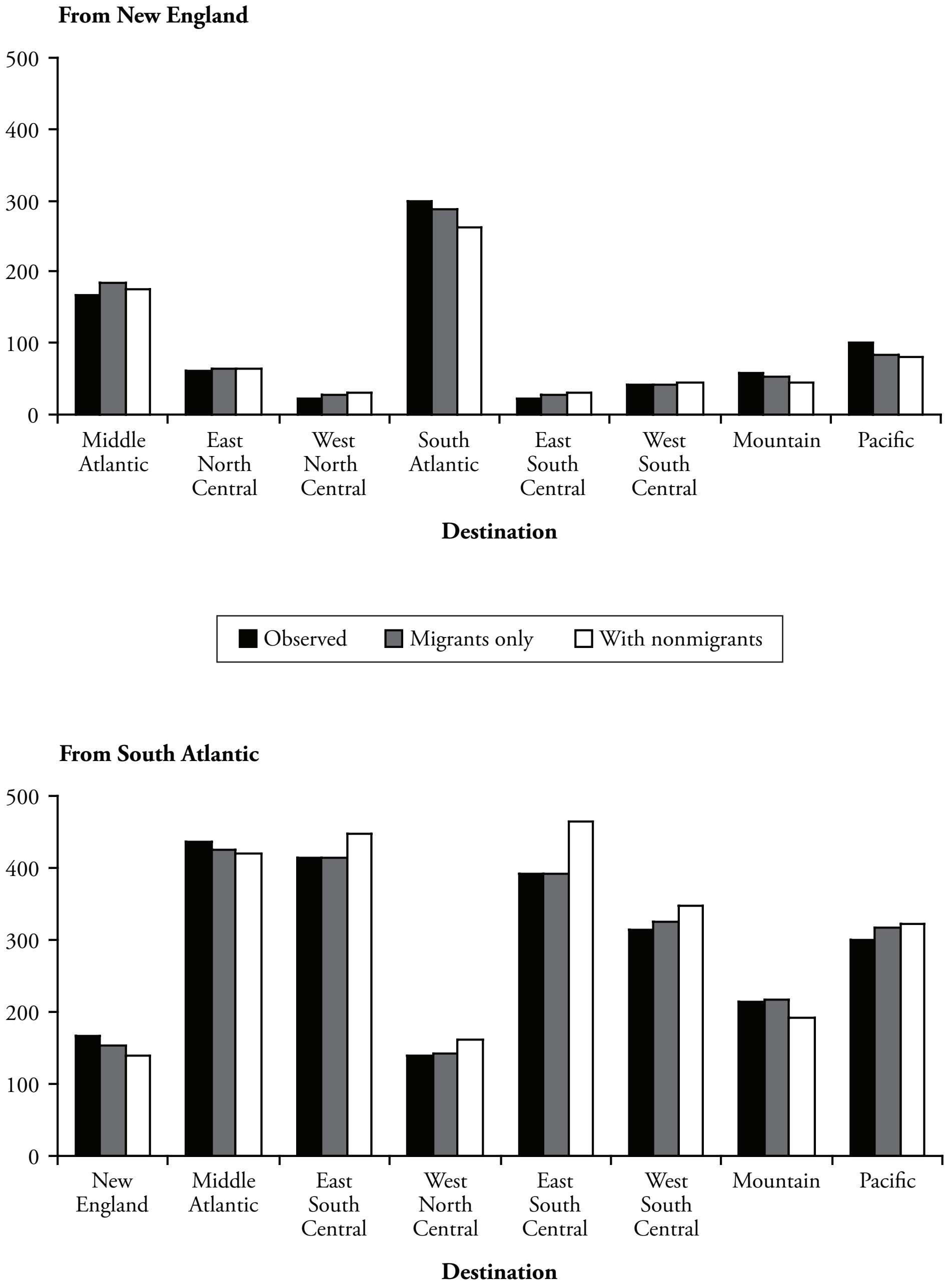
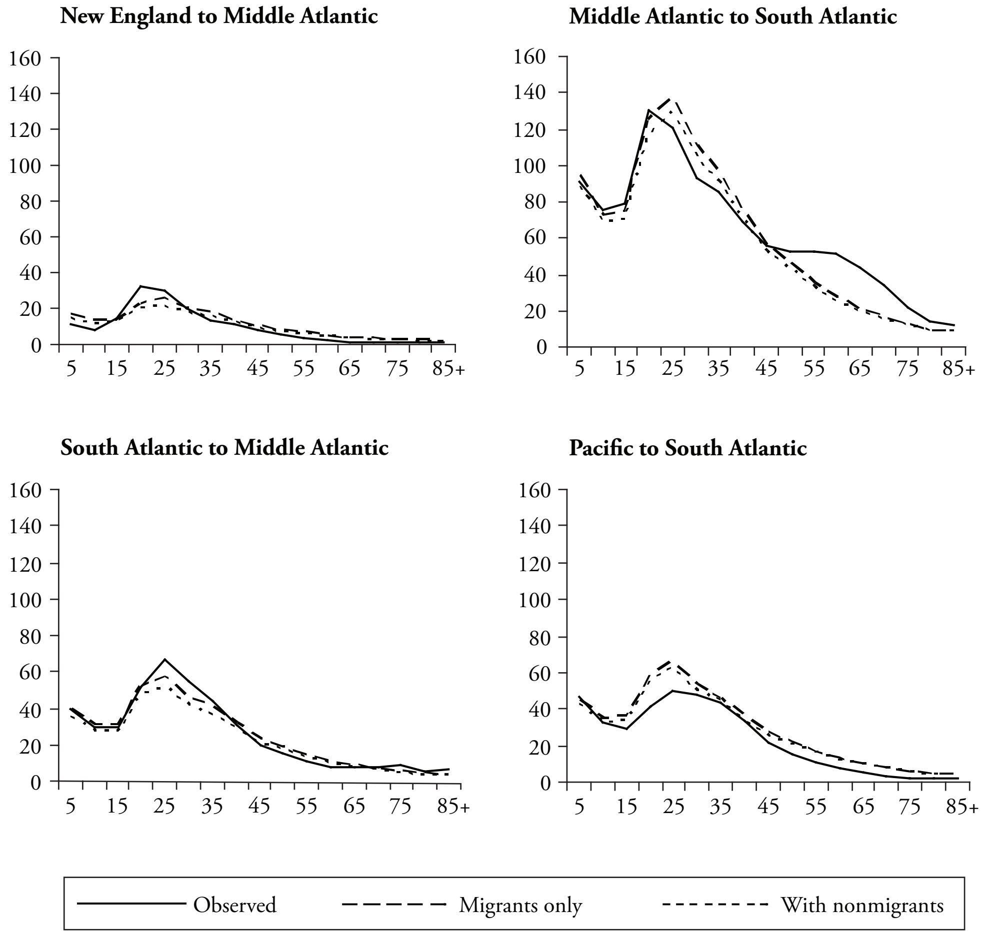


Figure 9. A Comparison of Infant Migration Log-Linear Model Predictions: Selected Age-Specific Interregional Migration Flows (in thousands) in the United States, 1995–2000



migrants-only model once again did a better job, with an overall likelihood ratio statistic of 104,962 versus 150,888 for the migrants and nonmigrants model. The Border–North Central, North Central–Border, Central-South, and South-Central flows had R^2 values of 0.911, 0.988, 0.929, and 0.933, respectively. The likelihood ratio statistics for these flows were lower for all the flows in Figure 10, except the South-Central flow.

In applying our strategy of indirect estimation in countries such as Mexico, one encounters a potential flaw, pointed out by a reviewer of this paper: the population under 5 years of age is often undercounted in less-developed countries, places where our methodology would potentially offer the greatest benefit. Note that this is a significant concern only if the degree of underenumeration varies by region. Moreover, the alternative option of using under-age-5 stocks that have been "corrected" to be consistent with assumed fertility

Table 6. The Spatial Structure of 0- to 4-Year-Old Birthplace-Specific Population Stocks in Mexico, 2000

	Destination				
Origin	Border	North Central	Central	South	Total
A. Observed Flows					
Border	0	26,511	15,549	2,072	44,132
North Central	41,664	0	21,826	4,052	67,542
Central	17,394	31,564	0	28,438	77,396
South	4,147	8,771	25,701	0	38,619
Total	63,205	66,846	63,076	34,562	227,689
B. Multiplicative Components					
Border	0.000	2.046	1.272	0.309	0.194
North Central	2.222	0.000	1.166	0.395	0.297
Central	0.810	1.389	0.000	2.421	0.340
South	0.387	0.774	2.402	0.000	0.170
Total	0.278	0.294	0.277	0.152	227,689

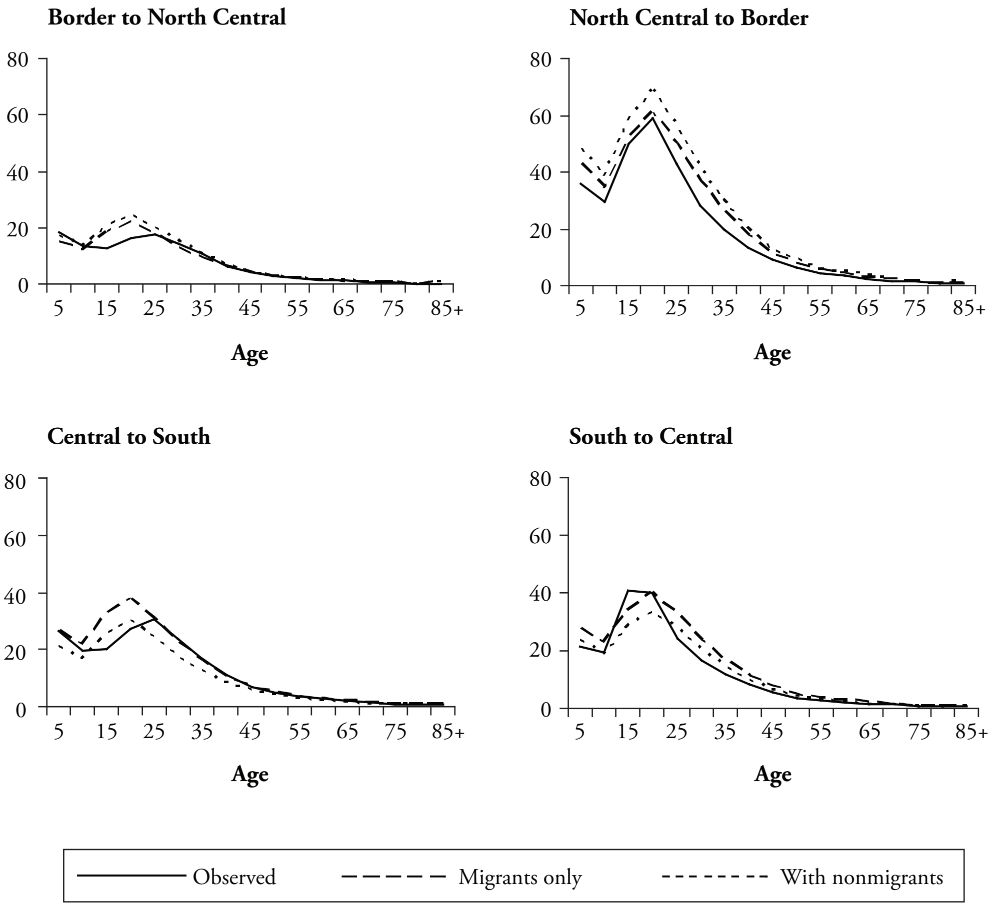
and mortality patterns in each region simply shifts the problem to a different potential flaw—that is, the need to specify the fertility and mortality patterns of migrant populations. Finally, a suggestion was made that a possibly better option would be to use the 5–9 years age group, despite the problem posed by multiple moves. We tried this option and came away with mixed results. We found that when the 5–9 birthplace-specific stocks of migrants were used in the offset (i.e., where structural zeros were inserted in the diagonal), the results were indeed somewhat better. But in that alternative, the migration flow marginal totals are assumed to be known. However, when 5–9 birthplace-specific population stocks were used in the offset (a more common situation), then the predicted flows of migrants were overestimated.

DISCUSSION AND CONCLUSION

The age structure of a population is a fundamental concept in demography, one that is normally depicted in the form of an age pyramid. The age structure of migration has also become a fundamental concept, one that can be expressed in the form of a model migration schedule (Rogers and Castro 1981). The spatial structure of an interregional system of origin-destination-specific migration streams, however, is a notion that lacks a widely accepted definition. In this article, we adopt the definition presented in Rogers, Willekens, Little, and Raymer (2002), which draws on the log-linear specification of the spatial interaction model (Willekens 1983)—a specification that involves a multicomponent breakdown of the matrix of flows under study. Such a formulation allows one to capture different features of a particular spatial structure of migration, with one set of parameters representing the effects of sizes of origin populations, another set representing the corresponding effects of the sizes of destination populations, and still another set representing the strengths of the linkages between these two populations.

The indirect estimation methods presented in this article assume the availability of some regional migration or population stock data to predict current (or future) migration flows. The use of migrants-only data yields more accurate results than the use of population

Figure 10. A Comparison of Infant Migration Log-Linear Model Predictions: Selected Age-Specific Interregional Migration Flows (in thousands) in Mexico, 1995–2000



data because the model then has structural zeroes in the diagonals and avoids the overwhelming influence carried by the otherwise nonzero diagonal elements representing the nonmigrants. But the improved accuracy comes at a cost: it needs an estimate of the nonmigrant populations that are subtracted from the marginal totals in order to obtain zeroes in the diagonals.

The size and age distribution of a particular migration stream are insufficient to characterize the flows of migrants; one also needs a description of the spatial interlinkages between origins and destinations. In certain instances, past patterns of migration from one region to another may be a better predictor of a current migration pattern than the particular characteristics of the two regions (Rogerson 1984). Where this is so, a method of estimation that uses a past spatial structure as part of its procedures is appropriate. Where this is not the case, the alternative initial estimates of the spatial linkages (interactions) between each pair of origins and destinations may be obtained from other auxiliary sources of information, for example, from the migration spatial structure exhibited by the under-5 population—one inferred from birthplace-specific residence data of that age group in a current census count (Rogers and Jordan 2004). The unique contribution of the log-linear modeling framework for the indirect estimation of migration is its ability to "discipline" these alternative initial estimates by imposing constraints on the estimated values—constraints that arise from associated historical data, partial data, or even qualitative or judgmental data (Rogers et al. 2003).

As we explained earlier, the U.S. Census Bureau is dropping its long-form questionnaire in 2010 and replacing it with a continuous monthly survey called the American Community Survey. This change provides more timely data, but the samples are smaller than have been provided by the decennial census, and the strategy of averaging accumulated samples over time mixes changing migration patterns. Moreover, the migration question refers to a one-year time interval instead of the five-year interval used since the 1960 census. For all of these reasons, it may be useful to have at hand a method for complementing or augmenting the collected data with indirect estimates of missing observations, particularly at fine levels of age, sex, and spatial disaggregation.

The migration data in less-developed countries, such as Mexico, can be even more problematic, making the log-linear framework presented in this article even more useful. However, certain hurdles posed by, for example, significant differential age misreporting and undercounting across regions, will need to be overcome. A National Academy of Sciences report on age-selective underenumeration concluded that, "Although age misreporting and selective underenumeration will continue to plague demographic studies, the recent evidence suggests that we can do a much better job of adjusting data for misreporting errors and of developing techniques for estimating fertility and mortality that are less sensitive to age reporting errors" (Ewbank 1981:87). The same can be said for the task of estimating migration.

In conclusion, the following observations need to be made. First, the multiplicative component model is a flexible and powerful framework for *analyzing* migration flows. Second, the log-linear model is an equally flexible and powerful framework for *estimating* migration flows. Third, estimation of migrant counts alone (with structural zeroes entered in the diagonals) yields more accurate estimates than does the corresponding migrants-plus-nonmigrants estimation procedure. Finally, future work should be directed at the potential improvements provided by the introduction of covariates in the statistical estimation process, for example, the association between the age composition of a population and that of its out-migrants (Little and Rogers 2007).

REFERENCES

Agresti, A. 2002. *Categorical Data Analysis*. Hoboken, NJ: John Wiley and Sons.

Ahmed, B. and J.G. Robinson. 1994. "Estimates of Emigration of the Foreign-Born Population: 1980–1990." Technical Working Paper No. 9. Population Division, U.S. Bureau of the Census, Washington, DC.

Bogue, D.J. 1969. *Principles of Demography*. New York: John Wiley and Sons.

Brass, W. 1974. "Perspectives in Population Prediction: Illustrated by the Statistics of England and Wales." *Journal of the Royal Statistical Society A* 137:532–70.

Ewbank, D.C. 1981. *Age Misreporting and Age-Selective Underenumeration: Sources, Patterns, and Consequences for Demographic Analysis*. Washington, DC: National Academy Press.

Heligman, L. and J.H. Pollard. 1980. "The Age Pattern of Mortality." *Journal of the Institute of Actuaries* 107(434):49–80.

Hill, K. 1985. "Indirect Approaches to Assessing Stocks and Flows of Migrants." Pp. 205–24 in *Immigration Statistics: A Story of Neglect,* edited by D.B. Levine, K. Hill, and R. Warren. Washington, DC: National Academy Press.

Little, J.S. and A. Rogers. 2007. "What Can the Age Composition of a Population Tell Us About the Age Composition of Its Outmigrants?" *Population, Space and Place* 13:23–39.

Nair, P.S. 1985. "Estimation of Period-Specific Gross Migration Flows From Limited Data: Bi-Proportional Adjustment Approach." *Demography* 22:133–42.

Plane, D.A. 1981. "Estimation of Place-to-Place Migration Flows from Net Migration Totals: A Minimum Information Approach." *International Regional Science Review* 6(1):33–51.

———. 1982. "An Information Theoretic Approach to the Estimation of Migration Flows." *Journal of Regional Science* 22:441–56.

Preston, S.H., P. Heuveline, and M. Guillot. 2000. *Demography: Measuring and Modeling Population Processes*. New York: Blackwell.

Raymer, J. Forthcoming. "Obtaining an Overall Picture of Population Movement in the European Union." In *The Estimation of International Migration in Europe: Issues, Models and Assessment,* edited by J. Raymer and F. Willekens. Chicester, United Kingdom: John Wiley and Sons.

Raymer, J., A. Bonaguidi, and A. Valentini. 2006. "Describing and Projecting the Age and Spatial Structures of Interregional Migration in Italy." *Population, Space and Place* 12:371–88.

Rogers, A. and L.J. Castro. 1981. "Model Migration Schedules." RR-81-30, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Rogers, A. and L. Jordan. 2004. "Estimating Migration Flows From Birthplace-Specific Population Stocks of Infants." *Geographical Analysis* 36(1):38–53.

Rogers, A. and J.S. Little. 1994. "Parameterizing Age Patterns of Demographic Rates With the Multiexponential Model Schedule." *Mathematical Population Studies* 4(3):175–94.

Rogers, A., F.J. Willekens, J.S. Little, and J. Raymer. 2002. "Describing Migration Spatial Structure." *Papers in Regional Science* 81:29–48.

Rogers, A., F.J. Willekens and J. Raymer. 2001. "Modeling Interregional Migration Flows: Continuity and Change." *Mathematical Population Studies* 9:231–63.

———. 2002. "Capturing the Age and Spatial Structures of Migration." *Environment and Planning A* 34:341–59.

———. 2003. "Imposing Age and Spatial Structures on Inadequate Migration-Flow Datasets." *The Professional Geographer* 55(1):56–69.

Rogerson, P.A. 1984. "New Directions in the Modelling of Interregional Migration." *Economic Geography* 60(2):111–21.

Rowland, D.T. 2003. *Demographic Methods and Concepts*. Oxford: Oxford University Press.

Schmertmann, C.P. 1992. "Estimation of Historical Migration Rates From a Single Census: Interregional Migration in Brazil 1900–1980. *Population Studies* 46(1):103–20.

Siegel, J.S. and D.A. Swanson, Eds. 2004. *The Methods and Materials of Demography*. Amsterdam: Elsevier Academic Press.

Snickars, F. and J.W. Weibull. 1977. "A Minimum Information Principle: Theory and Practice." *Regional Science and Urban Economics* 7:137–68.

Tobler, W. 1995. "Migration: Ravenstein, Thornthwaite, and Beyond." *Urban Geography* 16: 327–43.

United Nations. 1983. *Manual X: Indirect Techniques for Demographic Estimation*. New York: Department of International Economic and Social Affairs.

Warren, R. and E.P. Kraly. 1985. "The Elusive Exodus: Emigration From the United States." *Population Trends and Public Policy Occasional Paper*, No. 8. Population Reference Bureau, Washington, DC.

Warren, R. and J.M. Peck. 1980. "Foreign-Born Emigration From the United States: 1960 to 1970." *Demography* 17:71–84.

Willekens, F.J. 1980. "Entropy, Multiproportional Adjustment and the Analysis of Contingency Tables." *Systemi Urbani* 2/3:171–201.

———. 1982. "Multidimensional Population Analysis With Incomplete Data." Pp. 43–111 in *Multidimensional Mathematical Demography,* edited by K. Land and A. Rogers. New York: Academic Press.

———. 1983. “Log-Linear Modelling of Spatial Interaction.” *Papers of the Regional Science Association* 52:187–205.

———. 1999. “Modeling Approaches to the Indirect Estimation of Migration Flows: From Entropy to EM.” *Mathematical Population Studies* 7:239–78.

Zaba, B. 1987. “The Indirect Estimation of Migration: A Critical Review.” *International Migration Review* 21:1395–445.